Clinical Medicine 2011, Vol 11, No 3: 227–30

How to write high-quality questions for a high-stakes examination

Charles Twort and John Mucklow

On behalf of the Federation of Royal Colleges of Physicians of the UK, MRCP(UK) develops and delivers membership and specialty examinations that are recognised around the world as quality benchmarks. The following four articles highlight developments in the MRCP(UK) Diploma and specialty certificate examinations (SCEs). In the first article, Charles Twort and John Mucklow describe the process of ensuring questions for the written examinations meet rigorous academic standards. In the second article, Andrew Elder and colleagues summarise recent changes to PACES, which aim to ensure that passing candidates are competent across the range of clinical skills assessed. In the third article, John Mucklow describes the rationale for, and implementation of, the SCEs, which are now maturing past the launch phase and becoming an established part of the academic calendar. The final article in this series looks at the problem posed by cheating in the written exams. Trudie Roberts and colleagues outline the measures adopted by MRCP(UK) to remove opportunities for cheating and improve detection. We hope you enjoy these updates. Our aim, as always, is to provide examinations that are up-to-date and fit for purpose.

Jane Dacre

Medical director, MRCP(UK)

ABSTRACT – Both the MRCP(UK) written examinations and the specialty certificate examinations (SCEs) use single-best-answer questions to assess the knowledge and problem-solving skills of physicians in training. Since 1999, specialists and senior trainees have created large banks of questions for these examinations that cover the relevant curricula. Question-writing workshops provide detailed guidance on the design of questions that discriminate between candidates of differing ability, in a format and a style that aid speed reading and comprehension. Each question drafted is subjected to face-to-face peer review and subsequent stages of academic scrutiny before reaching the question bank, and later during the exacting processes of question selection and standard setting. Feedback to question writers at every level of scrutiny helps to support the development of question-writing competence, and the analysis of individual question performance provides some insight into optimal question design.

KEY WORDS: core medical training, one-best-of-five questions, postgraduate training, single-best-answer questions, specialist training

Introduction

It is over 10 years since the Federation of the Royal Colleges of Physicians of the UK decided to introduce single-best-answer (best-of-five) questions into the MRCP(UK) written examinations. It had become apparent that the multiple true/false format previously used in the Part 1 examination was being superseded internationally by the single-best-answer question and the decision was taken to follow this trend, and at the same time to introduce similar questions into the Part 2 written examination. This involved a considerable change in the question-writing process to build the necessary question bank. Small teams (specialty question groups; SQGs) of six invited question writers in each medical specialty (two from each medical college) were convened in 1999 and began to develop suitable questions. During the next three years, banks of single-best-answer questions were built for both the Part 1 and the Part 2 written examinations, and became the norm in diets from 2002 onwards.

By 2003, the initial SQGs had dwindled in number and were beginning to develop question writers' fatigue. Nomination of new members had not met the need for succession planning and question output began to fall. In 2004, the decision was made to advertise nationally for new members, and a gratifying response allowed for a doubling of team size. This coincided with the introduction of workshops to train new members, accelerating their development of the necessary competence, and question output grew apace – a further 2,000 questions have been added to the question bank every year since this time. Production of a question-writing manual that is updated annually established an agreed house style and non-medical editors were appointed to ensure that question format and appearance conformed to this.

Having established a bank of questions capable of sustaining the written examinations during the foreseeable future, it has been possible to review the process and consider ways of adapting it to future requirements. It seemed a good time to describe the process in some detail for those engaged in similar activities and to let all interested parties know exactly what it takes to produce a best-of-five question for the MRCP(UK) and the specialty certificate examinations (SCEs). The reader is referred to a companion article in this series for information on the development of the SCEs (pp 235–8).

The examinations

The focus of the MRCP(UK) Part 1 examination, which is aimed at those who have been qualified for 12–18 months, is to assess competence in relation to:

Charles Twort, chair, MRCP(UK) Specialty Question Group; **John Mucklow,** associate medical director for written examinations

MRCP(UK) Central Office, Royal College of Physicians, London

- core medical knowledge
- basic clinical problem-solving skills
- the management of common and important medical emergencies.

The Part 1 examination comprises two papers of 100 single-best-answer (best-of-five) questions.

By contrast, the focus of the MRCP(UK) Part 2 written examination, which is aimed at those who have occupied a core medical training post for 12–18 months, is to assess competence in relation to:

- advanced problem-solving skills
- the investigation, diagnosis, management and prognosis of a range of acute and chronic medical conditions
- the recognition and management of complications of a primary diagnosis and of medical therapies
- evidence-based medicine and nationally/specialty-based approved management guidelines.

The Part 2 written examination comprises three papers of 90 single-best-answer (best-of-five) questions.

Part 1 questions tend to be shorter than those in Part 2. The scenarios used in Part 1 questions have a maximum of 70 words, whereas those in Part 2 can be up to 300 words, depending on the number of investigation results presented and whether there is an associated image to interpret.

The principles of question writing

Multiple-choice questions come in different guises. The traditional multiple true/false type that asks candidates to identify all the correct statements listed can assess knowledge and comprehension, and is valuable for self-assessment as a guide to gaps in one's knowledge, but is limited to these objectives. By contrast, the single-best-answer question, which asks candidates to choose the best answer from, say, five plausible alternatives, assesses not only knowledge and comprehension but also the application of this knowledge to the synthesis of deductions from several assorted pieces of information, and the evaluation of laboratory and other numerical data.[1] In short, it can assess the problem-solving skills vital to the practising clinician.

Single-best-answer questions used in Federation examinations comprise:

- a stem (a clinical scenario, perhaps with some investigation results) that contains all the information necessary to anticipate the options, if not the correct answer; it should be complete, concise, unambiguous and free of any extraneous detail
- a lead-in question that ends with a question mark and asks the candidate to make a choice (eg, what is the most likely diagnosis? What is the most appropriate investigation?)
- five options that are grammatically compatible with the lead-in question, balanced in length and content, and usually in the same domain (diagnosis, investigation, management strategy).

A random choice of option will allow the uninformed candidates to score 20% and it is important that their ability to guess is not raised inadvertently above this level by providing grammatical or logical cues to the correct answer.

Answering 100 questions in three hours allows an average of under two minutes to read a question and select an answer. Candidates need to speed-read in order to save extra time for questions that require more thought. This is facilitated by arranging the clinical information (history, examination findings, listed investigations) in a strict order, as succinctly as possible, as well as presenting it in a way that does not disadvantage dyslexic candidates or those whose first language is not English.

The training workshops, which all question writers are obliged to attend as a prelude to their involvement in the process, explain the obligations of membership as well as the principles, the practice and the pitfalls of drafting single-best-answer questions. They offer guidance on choosing question topics and use of the same stem to introduce a variety of lead-in questions. The most suitable questions target areas of uncertainty, misunderstanding or unexpected ignorance among those trainees preparing for the examination who are as yet not confident of passing.

The first step in drafting a best-of-five question is to identify a suitably important topic, decide what precise knowledge one wishes to test and limit the question to a single testing point. Domains that might be tested include knowledge of basic science, knowledge of symptoms and signs, diagnosis, interpretation of results, choice of investigation and acute or chronic management. The next step is to write down the correct answer and identify four plausible alternative options that are either incorrect or clearly less correct than the correct answer, and are not widely divergent. The third step is to build the stem as a fictional scenario that includes all the information necessary to choose the correct answer (and no more). The final step is to check that the stem is written as economically as possible, and that the lead-in question and the options follow logically from it.

The process

MRCP(UK)

From the outset in 1999, SQGs met twice a year for two successive days to discuss drafted questions (12 per member) that had been submitted in advance and circulated electronically to members of each relevant group. The chair of each group was either a generalist or a specialist in a different clinical discipline, to ensure that subject matter did not become too arcane. In the first instance, editing at the meeting was done on paper copies, but a laptop computer, electronic projector and screen are now provided for each group so that all those present can view the editing process. The work rate increased substantially after the appointment of non-medical editors allowed for amendments to format and style to be made before each meeting and leaving discussion to focus on the medical rather than the presentational aspects of each question. SQGs are required to discuss at

least eight questions an hour and an experienced group can process and agree over 100 questions during a two-day meeting. If the group is large enough to subdivide (subgroups must comprise at least four experienced members to ensure a critical level of academic scrutiny) the output can be multiplied by the number of subgroups.

Face-to-face peer review meetings of this type (which alternate between England and Scotland) are expensive to hold, as they involve substantial travel and accommodation costs. However, experience suggests that the quality of scrutiny that can be brought to bear on the question material justifies this outlay and affords considerable confidence in the quality of the product.

The SQG membership comprises chiefly registered specialists but also a substantial number of specialty trainees who have passed MRCP(UK). There are now 16 separate groups (Table 1). The tenure of membership is normally five years. After this time, members either stand down or become chairs of other groups. Those engaged in the work receive no tangible reward for their efforts, but enjoy returning for the intellectual stimulation the meeting affords. Having to defend one's drafted questions before one's peers means that each must be thought through in advance and based on evidence, which forces one to challenge one's assumptions. Most would agree that the quality of continuing professional development (CPD) that the meetings offer is second to none.

Specialty certificate examinations

The process of question writing and scrutiny adopted for the SCEs was almost identical to that developed for MRCP(UK). However, owing to the need to produce a sizeable question bank quickly, there were differences of scale. Each specialty recruited up to 25 registered specialists, who attended one of a series of workshops before convening as a question-writing group (QWG). Wherever possible, specialists with experience of writing questions for MRCP(UK) were invited to transfer to provide role models for those writing questions for the first time. QWG members were asked to draft 15 questions before each meeting for editing. This requirement, together with the size of the QWGs, allowed them to subdivide into up to four subgroups, which meant that a specialty could process over 300 questions per meeting. At each peer-review meeting, subgroups are chaired by an experienced question writer in the specialty. This scale of activity has allowed the generation of over 10,000 in the four years since question writing began. Once individual specialties have well over 1,000 questions, it will be possible to scale down the effort to match that of the SQGs. There are now 12 specialties for which the Federation has developed SCEs in collaboration with specialist societies (Table 2).

Table 1. MRCP(UK) specialty question groups.

Cardiology
Clinical haematology
Clinical pharmacology, therapeutics and toxicology
Clinical science
Dermatology
Endocrinology and diabetes mellitus
Gastroenterology
Geriatric medicine
Infectious diseases and genitourinary medicine
Medical oncology
Nephrology
Neurology
Ophthalmology
Psychiatry
Respiratory medicine
Rheumatology

The feedback

Following each SQG meeting, all questions are reviewed in detail by an experienced member of the relevant examining board before they are entered into the question bank, and fewer than 10% are returned for further work. Following the SCE QWG meetings there is no extra layer of academic scrutiny in advance of question review by the relevant specialty examining board, although the questions are reviewed by the associate medical director and occasional items are returned, usually on the grounds that they are not single-best-answer questions, or that the options are not adequately supported by information in the stem.

Further insight into the suitability of questions is provided by the analysis of question performance after each diet. This compares the performance of five quintiles of the examination cohort in relation to each question. A good quality question should be answered correctly by 35–85% of just-passing candidates (defined as those scoring an overall mark within 10% of the pass mark). There should also be an obvious positive correlation between the performance of the cohort on the individual question and in the examination as

Table 2. Specialty certificate examination (SCE) question writing groups.

Acute medicine
Dermatology
Endocrinology and diabetes mellitus
Gastroenterology
Geriatric medicine
Infectious diseases
Medical oncology
Nephrology
Neurology
Palliative medicine (SCE due for introduction in 2011)
Respiratory medicine
Rheumatology

a whole (ie the question should be answered correctly by appreciably more passing candidates than failing candidates). A reasonable proportion of candidates (especially those who did not pass) should also have chosen each incorrect option.

Conclusion

Writing high-quality single-best-answer questions for high-stakes examinations requires training and experience. The process must be supported by critical peer review to ensure that the question content is accurate and the challenge appropriate to the candidates' ability. Most writers require a workshop and at least two peer-review sessions before they start to feel confident. The process is highly rewarding and contributes to the CPD of all those involved. The process underpins the validity and reliability of the MRCP(UK) written examinations and the SCEs.

Acknowledgements

We acknowledge the enormous contributions of individual members of the SQGs and QWGs, past and present, in developing the question banks, of the SQG Part 1 and Part 2 MRCP(UK) medical secretaries in vetting and editing the questions, and of the editorial and secretarial staff in the MRCP(UK) Central Office in preparing for each peer-review process and collating its outcome.

Reference

1 Case SM, Swanson DB. *Constructing written test questions for the basic and clinical sciences*, 3rd edn. Washington, DC: National Board of Medical Examiners, 2002. www.nbme.org/publications/item-writing-manual.html

Address for correspondence: Dr C Twort, MRCP(UK) Central Office, 11 St Andrews Place, Regent's Park, London NW1 4LE.
Email: charles.twort@gstt.nhs.uk